

DEVELOPMENT OF GAS DENSITY MEASUREMENT MODEL (TANK) BASED  
ON PRINCIPAL COMPONENT ANALYSIS (PCA) TECHNIQUE

WAN MOHD HIDAYAT BIN WAN ABDULLAH

A thesis submitted in fulfillment  
of the requirements for the award of the degree of  
Bachelor of Chemical Engineering

Faculty of Chemical & Natural Resources Engineering  
University College of Engineering & Technology Malaysia

November 2006

## DECLARATION

I declare that this thesis entitled “*Development Gas Density Measurement Model (Tank) Based on Principal Component Analysis (PCA) Techniques*” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....  
Name : WAN MOHD HIDAYAT BIN WAN ABDULLAH  
Date : 22<sup>nd</sup> NOVEMBER 2006

*To my beloved mother and father...  
Thank you for bringing me onto this beautiful world  
My love...you are always in my mind*

## **ACKNOWLEDGEMENT**

As a vicegerent who is being created, I put myself down to express my highest thankful to Allah S.W.T for giving me strength and spirit to complete this final project with full of pride and dignity.

I would like to express my deepest gratitude to the following persons for their unlimited and kindness help as well as guidance enabled me to complete this research project in time as a partial fulfillment of the requirement of the degree of Bachelor Engineering (Chemical).

To Tuan Hj. Mohd Yusri B. Mohd Yunus, my lovely supervisor, thanks a lot for helping me during the progress of the research project, either in a way of searching the required information as well as guiding me to be a good person. Without his generosity in sparing his precious time to guide me and answer my doubts, the aim of project may not be fulfilled.

My highest appreciation to my much closed friends Mimiey, Faiss, Deqwan, Rufizal, and Anuar who inspired me a lot and in addition I am very grateful to my family members, no amount of gratitude could repay their kindness of being there as well as the patience to put up with my whim.

I also would like to thank my fellow course mates and friends for their assistances, advices and ideas in producing a resourceful, fruitful and practicable research. May God bless all of you.

## **ABSTRACT**

The real importance of controlling gas density is based on the fact that it contributes significantly to the composition of the gas being controlled. Even with small variations can cause significant process upsets if not adequately controlled due to changing in composition. This research purposely to develop a model of gas density measurement as the process instrument nowadays hardly to correlate simultaneously the input variables to the output variables. In this respect, density is known as output whereby it depends on other variables (input) which are pressure and temperature of the given process. In other words, all variables which can be directly measured and contribute to the primary variables are known as 'secondary variables'. Thus, the Principal Component Analysis (PCA) approach is used plus the MATLAB programming to ensure the gas density measurement model is valid as the real process. At the end of this research, the developed model can be used to identify the best configurations of inputs variables in producing the specified output that intended. The best outcome is the effect of temperature and pressure to the density is very significant as the relationships between those variables are proportional to each other.

## ABSTRAK

Kepentingan utama dalam pengawalan ketumpatan gas adalah menyangkan secara signifikan komposisi gas yang dikawal. Walau dengan perbezaan yang kecil, ianya boleh menyebabkan proses tidak menepati seperti yang dikawal melihat kepada perubahan komposisi. Kajian ini bertujuan membentuk sebuah model pengiraan ketumpatan gas seperti kerana seperti yang kita ketahui, proses instrumen hari ini sukar untuk menghubungkan secara serentak pembolehubah masuk terhadap pembolehubah keluar dalam satu sesi proses yang sama. Dalam kaian ini, ketumpatan adalah pembolehubah keluar manakala suhu dan tekanan adalah pembolehubah masuk. Dalam erti kata yang lain, pembolehubah yang boleh dikira secara langsung merupakan pembolehubah sekunder. Oleh itu, pendekatan Analisis Prinsip Komponen (PCA) bersama simulasi MATLAB digunapakai untuk memastikan model pengiraan ketumpatan gas boleh memberikan bacaan seperti proses sebenar. Kepelbagaian model yang sudah dimajukan atau diperkembangkan pada akhir projek ini boleh digunakan untuk mengenalpasti konfigurasi pembolehubah keluar seperti yang dikehendaki. Keputusan menunjukkan, pembolehubah keluar bergantung secara langsung dengan pembolehubah masuk kerana hubungan di antara pembolehubah masuk dan pembolehubah keluar adalah berkadaran secara langsung.

**TABLES OF CONTENTS**

<b>CHAPTER</b>	<b>TITLE</b>	<b>PAGE</b>
	<b>TITLE PAGE</b>	<b>i</b>
	<b>DECLARATION</b>	<b>ii</b>
	<b>DEDICATION</b>	<b>iii</b>
	<b>ACKNOWLEDGEMENT</b>	<b>iv</b>
	<b>ABSTRACT</b>	<b>v</b>
	<b>ABSTRAK</b>	<b>vi</b>
	<b>TABLES OF CONTENT</b>	<b>vii</b>
	<b>LIST OF TABLES</b>	<b>x</b>
	<b>LIST OF FIGURES</b>	<b>xi</b>
	<b>LIST OF ABBREVIATION OR NOMENCLATURE</b>	<b>xii</b>
	<b>LIST OF APPENDICES</b>	<b>xiii</b>

<b>I</b>	<b>INTRODUCTION</b>	
1.0	Overview and Fundamental of Ideal Gas Law	1
1.0.1	Ideal Gas Law with Constraints	2
1.1	Problem Statement	4
1.2	Objectives	5
1.3	Scopes of Study	5
1.4	Research Contribution	5
1.5	Summary of Report	6
<b>II</b>	<b>FUNDAMENTAL THEORY</b>	
2.0	Introduction	7
2.1	Gas Density Measurement	9
2.1.1	Composition of Air	10
2.2	Principal Component Analysis (PCA)	11
2.2.1	Original Data Matrix, Variance-Covariance Matrix and Correlation Matrix	16
2.2.2	Determine the Eigenvalue and Eigenvector Of the Matrix	19
2.2.3	Number of Principal Component Required	20
<b>II</b>	<b>LITERATURE RIVIEW</b>	
3.0	Introduction	24
3.1	Gas Density Measurement	24
3.2	Principal Component Analysis (PCA)	29
<b>IV</b>	<b>METHODOLOGY</b>	
4.0	Introduction	35
4.1	Case Study of AFPT 921	35
4.2	Methodology Framework	37



	4.2.1 Data Gathering	38
	4.2.2 Density Calculation	38
	4.2.3 Plot Profile	38
	4.2.4 Model Development	38
	4.2.5 Validation	39
<b>V</b>	<b>RESULT AND DISCUSSION</b>	
	5.0 Data Gathering	40
	5.1 Gas Density Measurement	43
	5.2 Plot Profile	46
	5.3 Development of Principal Component Analysis Model	47
	5.4 The PCA Codes	49
	5.5 Validation	51
	5.6 Analysis and Discussion	54
<b>VI</b>	<b>CONCLUSION</b>	
	6.0 Conclusions	55
	6.1 Recommendation	55
	<b>REFERENCES</b>	<b>56</b>
	<b>APPENDIX I</b>	<b>59</b>
	<b>APPENDIX II</b>	<b>63</b>

**LIST OF TABLES**

<b>TABLES NO.</b>	<b>TITLE</b>	<b>PAGE</b>
2.1	The Composition of Dry Air is Relatively Uniform Around the World.	10
5.1	Temperatures in Pipeline (Experiment Data)	37
5.2	Pressures in Pipeline (Experiment Data)	42
5.3	Results of Density Calculation	45
5.4	Results of Gas Density from Programming	51
5.5	Comparison between Manual and Simulation	53

**LIST OF FIGURES**

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
1.1	PvT Surface	3
2.1	PCA Graphical matrix representation	13
2.2	System for Two Variables Distribution	15
2.3	Graphical Representation for PCA	15
2.4	Scree Plot	22
2.5	Principal Component Analysis Procedures	23
4.1	Flowcharts for Methodology	37
5.1	Profile Density versus Temperature	46
5.2	Profile Density versus Pressure	47
5.3	The Programming Data Set in M-file	48
5.4	The Programming Data Set in M-file (contd)	48

## LIST OF ABBREVIATION OR NOMENCLATURE

T	temperature in pipeline, °C
P	pressure in pipeline, Pa
$\rho$	density in tank, kg/m <sup>3</sup>
X	data matrix, consisting of the set of all data vectors, one vector per column
N	the number of column vectors in the data set
M	the number of elements in each column vector
$\mu$	vector of empirical means, one mean for each row $m$ of the data matrix
s	vector of empirical standard deviations, one standard deviation for each row $m$ of the data matrix
Z	z-scores, computed using the mean and standard deviation for each row $m$ of the data matrix
C	covariance matrix
R	correlation matrix
V	matrix consisting of the set of all eigenvectors of <b>C</b> , one eigenvector per Column
W	matrix of basis vectors, one vector per column, where each basis vector is one of the eigenvectors of <b>C</b> , and where the vectors in <b>W</b> are a sub-set of those in <b>V</b>
Y	matrix consisting of $N$ column vectors, where each vector is the projection of the corresponding data vector from matrix <b>X</b> onto the basis vectors contained in the columns of matrix <b>W</b> .
Cov	Covariance
Eig	Eigenvalue

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
<b>I</b>	<b>CASE STUDE OF AIR FLOW PRESSURE TEMPERATURE (AFPT921 PLANT)</b>	<b>59</b>
<b>II</b>	<b>CODING FOR GAS DENSITY MEASUREMENT MODEL USING PCA TECHNIQUE IN MATLAB</b>	<b>63</b>

## CHAPTER I

### INTRODUCTION

#### 1.0 Overview and Fundamental of Ideal Gas Law

An ideal gas is defined as one in which all collisions between atoms or molecules are perfectly elastic and in which there are no intermolecular attractive forces. One can visualize it as a collection of perfectly hard spheres which collide but which otherwise do not interact with each other. In such a gas, all the internal energy is in the form of kinetic energy and any change in internal energy is accompanied by a change in temperature.

An ideal gas can be characterized by three state variables: absolute pressure (P), volume (V), and absolute temperature (T). The relationship between them may be deduced from kinetic theory and is called the Ideal Gas Law

$$PV = nRT = NKT \quad (1.1)$$

Where,

$n$  = number of moles

$R$  = universal gas constant = 8.3145 J/mol K

$N$  = number of molecules

$k$  = Boltzmann constant =  $1.38066 \times 10^{-23}$  J/K =  $8.617385 \times 10^{-5}$  eV/K

$k = R/N_A$

$N_A$  = Avogadro's number =  $6.0221 \times 10^{23}$  /mol

The ideal gas law can be viewed as arising from the kinetic pressure of gas molecules colliding with the walls of a container in accordance with Newton's laws.

But there is also a statistical element in the determination of the average kinetic energy of those molecules. The temperature is taken to be proportional to this average kinetic energy; this invokes the idea of kinetic temperature. One mole of an ideal gas at STP occupies 22.4 liters.

### 1.0.1 Ideal Gas Law With Constraints

For the purpose of calculations, it is convenient to place the ideal gas law in the form:

$$\frac{P_i V_i}{T_i} = \frac{P_f V_f}{T_f} \quad (1.2)$$

where the subscripts i and f refer to the initial and final states of some process. If the temperature is constrained to be constant, this becomes:

$$P_i V_i = P_f V_f \quad (1.3)$$

which is referred to as Boyle's Law. If the pressure is constant, then the ideal gas law takes the form:

$$\frac{V_i}{T_i} = \frac{V_f}{T_f} \quad \text{or} \quad V_f \frac{T_f}{T_i} \quad (1.4)$$

which has been historically called Charles' Law. It is appropriate for experiments performed in the presence of a constant atmospheric pressure.

All the possible states of an ideal gas can be represented by a  $PvT$  surface as illustrated Figure 1.1. The behavior when any one of the three state variables is held constant is also shown.

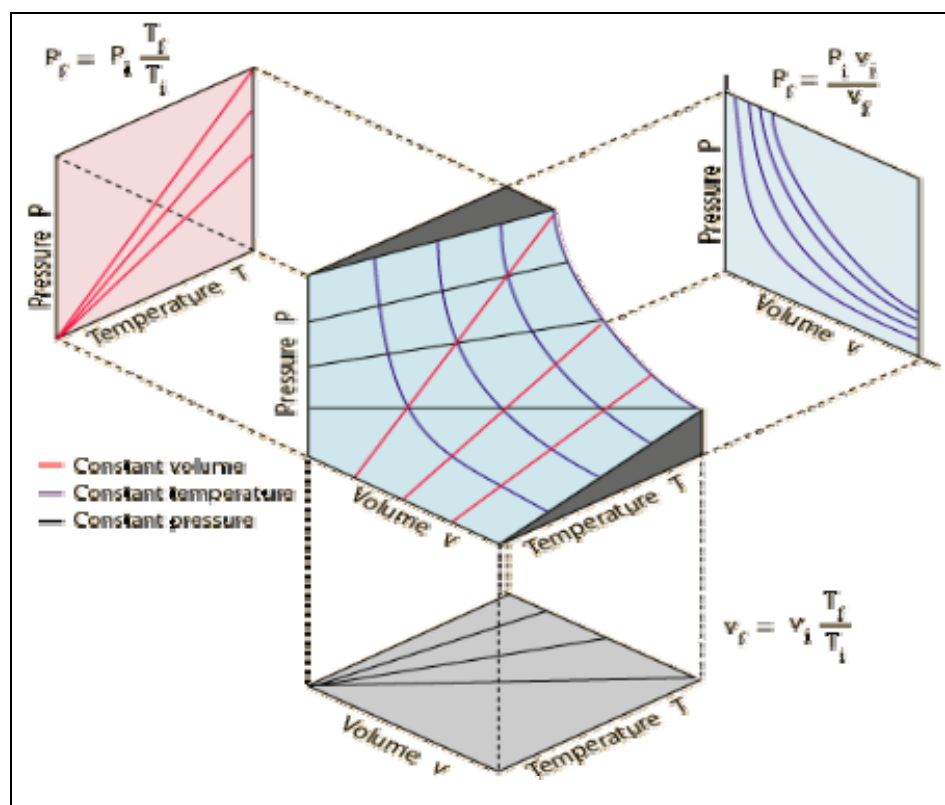


Figure 1.1 PvT Surface



## 1.1 Problem Statement

The real importance of controlling gas density is based on the fact that it contributes significantly to the composition of the gas being controlled. Even with small variations can cause significant process upsets if not adequately controlled due to changing in composition. As for example, the variations in gas composition experienced resulted in the production engineers using a higher oxygen level than was actually needed to allow safety margin. As a result, this will lead to lower process efficiency and higher production cost [1].

However, density cannot be measured directly from the process due to several reasons as the followings [2].

- a) Lack of appropriate on-line instrumentation.
  - i) Process operation has to depend on laboratory analysis.
  - ii) The cost of installing a sensor may not be justified.
- b) Reliability of on-line instruments.
  - i) Slow feedback.
  - ii) Lack of measurement in a timely manner.

In this respect, density is known as 'primary variables' whereby it depends on other variables (secondary variables) which are pressure and temperature of the given process. In other words, all variables which can be directly measured and contribute to the primary variables are known as 'secondary variables'. As a result, the quality of the product cannot be measured and analyzed in on-line or real time manner and this could lead to off-specification products.

Besides that, those primary variables are also not being monitored and controlled thoroughly as well as simultaneously with all the main secondary variables that available in the plant. Eventually, the analysis being carried out is lacking of reliability aspect, in the sense that it does not incorporate all secondary variables that potentially affect upon the primary variables when changes occur.

## **1.2 Objectives**

- i. The aims of this study is to model the relationships between quality output (density) and secondary outputs (temperature and pressure) using Principle Component Analysis (PCA).
- ii. To develop a program to measure the density from AFPT 921 plant.

## **1.3 Scopes of Study**

There are five major scopes that have been identified in order to achieve the objectives of this research:

- i) Data gathering from Air-Flow-Pressure-Temperature (AFPT) Plant to obtain a set of secondary outputs data.
- ii) Calculate the primary output by using Microsoft Excel.
- iii) To develop variable profile
- iv) Modeling the gas density measurement by using PCA.
- v) Comparing performance result of the gas density measurement modeled with the conventional measurement available.

## **1.4 Research Contribution**

The main outputs of this research are:

- i) A new gas density measurement model is proposed and developed by using PCA.
- ii) A new programming on gas density measurement is developed.

## **1.5 Summary of Report**

This thesis contains five chapters. The first chapter comprises of the introduction of the research, fundamental of Ideal Gas Law, objectives of the research, scope of the research and contributions of this research. Chapter two reviews the fundamental theory of gas density measurement and Principal Component Analysis (PCA).

Chapter three consists of the literature review which is reveals of the previous research and study about gas density measurement and the Principal Component Analysis (PCA)

Chapter four expresses of the research methodology which is in this chapter, the case study on AFPT 921 plant has been made while method algorithm is developed as a guidance to complete the research.

Chapter five presents the result and discussion of the research. The results are systematically presented into the tables and graph togetherness with the discussion and justification.

## CHAPTER II

### FUNDAMENTAL THEORY

#### 2.0 Introduction

Model AFPT 921 is a process control training system that uses only air to simulate gas, vapor or steam. Air is readily available from a compressor. It provides the simple gas physical processes where the measurement and control of their important variables of flow, temperature and pressure can be studied.

In general, density is one of gas physical properties in which it is very much dependant on the pressure and temperature of the given process under consideration. This is very true as illustrated in the Perfect Gas Law as follows:

$$d = \frac{m}{V} = \frac{P}{T} \cdot \frac{M}{R} = \frac{P}{T} \times \text{constant} \quad (2.1)$$

Whereby,

P = Pressure

V = Volume

m = Mass

M = Molecular weight

R = 8.314 J/mol.K

T = Temperature

$$\text{Therefore, } d = \frac{m}{V} = \frac{P}{T} \cdot \frac{M}{R} = \frac{P}{T} \times \text{constant} \quad (2.2)$$

for any given gas. For instance, the following shows the real issue in measuring gas density.

Assume the total quantity of gas through a pipeline is held constant at 50 kg/hr but the flowing pressure and temperature can be varying. Initially, the flow rate is at the Normal Condition (NTP) of 0°C temperature and 1 atm pressure, and increase gradually to 120°C and 1.485 atm respectively. As a result, the actual volume of the gas has expanded as well as contracted due to temperature and pressure changes at the same time.

Whilst the initial gas density is 1.293 kg/Nm<sup>3</sup>, the new calculated density could be calculated approximately as follows:

Initial density = 1.293 kg/Nm<sup>3</sup> at 0°C (273°K) and 1 atm (NTP) .

$$\begin{aligned} \text{Therefore, new density} &= 1.293 \times \frac{273\text{K}}{393\text{K}} \times \frac{1.485 \text{ atm}}{1.0 \text{ atm}} \\ &= 1.334 \text{ kg/ m}^3 \text{ at } 120^\circ\text{C, } 1.485 \text{ atm} \end{aligned}$$

## 2.1 Gas Density Measurement

The gas density can be measured by many ways. Nowadays there are a lot of equipment were invented for the purpose mean. But when it comes to the basic element which is reached to the assuming of ideal gas, it is definitely has to reveal the characteristics of the ideal gas.

The ideal gas law can be manipulated to find the density of a substance, the manipulation of the ideal gas law as shown below:

$$PV=nRT \quad n=\frac{m}{mw}, \quad \rho=\frac{m}{V}$$

Rearrange the equation,

$$PV=\frac{m}{mw}RT \tag{2.3}$$

$$\frac{m}{V}=\frac{mw}{R}\left(\frac{P}{T}\right)$$

$$\rho = \text{constant} \left( \frac{P}{T} \right)$$

In several cases, the P and T is to be manipulated as the density very dependant of changing the pressure and temperature.

### 2.1.1 Composition of Air

Air is a mixture of gases with overall physical and chemical properties consistent with its composition. The "real" air that we encounter on a day-to-day basis, and use for breathing, burning fuels for heating, transportation, power generation and many other purposes, can be thought of as a mixture of three types of components:

- i) Standard dry air, which is mainly composed of three gases. They are nitrogen (about 78%), oxygen (about 21%), and argon (a bit less than 1%). All three of these gases can be economically recovered as industrial gas products. Standard dry air also contains a small amount of carbon dioxide, and very small amounts of neon, helium, krypton, hydrogen and xenon

**Table 2.1** The composition of dry air is relatively uniform around the world.

<b>Standard Dry Air Composition (Detailed Analysis):</b>					
<b>Gas</b>	<b>% by Volume</b>	<b>% by Weight</b>	<b>Parts per Million (V)</b>	<b>Chemical Symbol</b>	<b>Molecular Weight</b>
Nitrogen	78.08	75.47	780805	N <sub>2</sub>	28.01
Oxygen	20.95	23.20	209450	O <sub>2</sub>	32.00
Argon	0.93	1.28	9340	Ar	39.95
Carbon Dioxide	0.038	0.0590	380	CO <sub>2</sub>	44.01
Neon	0.0018	0.0012	18.21	Ne	20.18
Helium	0.0005	0.00007	5.24	He	4.00
Krypton	0.0001	0.0003	1.14	Kr	83.80
Hydrogen	0.00005	Negligible	0.50	H <sub>2</sub>	2.02
Xenon	8.7 x 10 <sup>-6</sup>	0.00004	0.087	Xe	131.30

## 2.2 Principal Component Analysis

Principal Components Analysis (Jackson, 1980; Geladi and Kowalski, 1986; Wold *et al.*, 1987; Jackson, 1991; Johnson and Wichern 1992) can be described via using mathematical representation and graphical representation. From the mathematical point of view, Principal Components Analysis is a multivariate technique, which can transform a set of original variables  $x_1, x_2, \dots, x_m$  to a set of new variables  $P_1, P_2, \dots, P_m$ . Actually, the newly formed variables are called Principal Components (PC), which are the linear combinations of the original variables. The mathematical representations that describe the transformation are shown as follow:

$$P_1 = v_{1,1}x_1 + v_{1,2}x_2 + \dots + v_{1,m}x_m$$

$$P_2 = v_{2,1}x_1 + v_{2,2}x_2 + \dots + v_{2,m}x_m$$

.

.

$$P_m = v_{m,1}x_1 + v_{m,2}x_2 + \dots + v_{m,m}x_m \quad (2.4)$$

$$P_{m \times n} = V_{m \times m} X_{m \times n}^T \quad (2.5)$$

The original matrix,  $X$  has  $m$  variables with each variable has  $n$  measurements. The data are arranged in the form of  $n \times m$ , where the measurements of a variable are organized in the form of column vector. A row vector, which contains the measurement of all variables is denoted as  $x_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$ , while  $[x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]^T$  is the transpose of the row vector.  $V$  is known as the *eigenvector* matrix.  $V$  is the weighting matrix in forming the linear combinations of the original variables. The *eigenvectors* matrix,  $V$ , contains *eigenvectors* or loading vectors  $v_1, v_2, \dots, v_m$ . Each *eigenvector*  $V_j$  is a row vector which contains the arrangement of elements as  $V_j = [v_{j,1} \ v_{j,2} \ \dots \ v_{j,m}]$ . The *eigenvectors* matrix can be shown in the following matrix:



$$\text{Eigenvectors matrix, } \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_m \end{bmatrix} = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,m} \\ v_{2,1} & v_{2,2} & \dots & v_{2,m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ v_{m,1} & v_{m,2} & \dots & v_{m,m} \end{bmatrix} \quad (2.6)$$

$\mathbf{P}$  is the principal components scores matrix. A particular principal component score,  $P_{i,j}$ , where  $j < m$ , is obtained via the product of a row vector variable measurement  $[x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$  to the particular *eigenvector*  $\mathbf{V_j}$ , which is shown in the following equation:

$$P_{i,j} = \mathbf{v}_j \mathbf{x}_i^T \quad (2.7)$$

$$P_{i,j} = \begin{bmatrix} v_{j,1} & v_{j,2} & \dots & v_{j,m} \end{bmatrix} \cdot \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \cdot \\ \cdot \\ x_{i,m} \end{bmatrix} \quad (2.8)$$

$$\mathbf{P}_{j(1 \times n)} = \mathbf{v}_{j(1 \times m)} \mathbf{X}_{(m \times n)}^T \quad (2.9)$$

$\mathbf{p_j}$  is the score vector for the Principal Component  $j$ , while  $\mathbf{v_j}$  is the *eigenvector*  $j$  in order to form Principal Component. Equation 2.8 is used to determine the principal component score matrix  $\mathbf{P}$ , which contains  $n$  scores for each principal component:

$$\mathbf{P}_{(m,n)} = \mathbf{V}_{(m \times m)} \mathbf{X}_{(m \times n)}^T \quad (2.10)$$

On the other hand, the original variables can also be represented in term of newly formed variables as follow and as shown in Figure 2.1:

$$\mathbf{X}_{(n \times m)} = \mathbf{P}_{1(n \times l)}^T \mathbf{V}_{1(l \times m)} + \mathbf{P}_{2(n \times l)}^T \mathbf{V}_{2(l \times m)} + \dots + \mathbf{P}_{m(n \times l)}^T \mathbf{V}_{m(l \times m)} \quad (2.11)$$

$$\mathbf{X}_{(n \times m)} = \mathbf{P}_{(n \times m)}^T \mathbf{V}_{(m \times m)} \quad (2.12)$$

$$\begin{aligned} \begin{bmatrix} \mathbf{X} \end{bmatrix}_n^m &= \begin{bmatrix} \mathbf{p}_1^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_1 \end{bmatrix}_1^m + \begin{bmatrix} \mathbf{p}_2^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_2 \end{bmatrix}_1^m + \dots + \begin{bmatrix} \mathbf{p}_m^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_m \end{bmatrix}_1^m \\ &= \begin{bmatrix} \mathbf{p}^T \end{bmatrix}_n^m \begin{bmatrix} \mathbf{v} \end{bmatrix}_m^m \end{aligned}$$

**Figure 2.1** PCA graphical matrix representation

Where $\mathbf{X}$	: original data matrix	dimension: $n \times m$
$\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_m$	: eigenvectors	dimension: $1 \times m$
$\mathbf{V}$	: eigenvector matrix	dimension: $m \times m$
$\mathbf{V}_a$	: matrix that contains a eigenvectors	dimension: $a \times m$
$\mathbf{P}_j$	: Principal Component j score vector	dimension: $1 \times n$
$\mathbf{P}$	: Principal Component score matrix	dimension: $m \times n$
$\mathbf{P}_a$	: Principal Component score matrix for PC I till PC a	dimension: $a \times n$

From Equation 2.1, the number of original variables,  $x_m$  equals to the number of the newly formed variables,  $P_m$ . If all Principal Components are used to represent the original variables, the original raw data matrix is reproduced as shown in Equation 2.9. If this is the condition, the purpose for using Principal Components Analysis as dimension reduction technique will be lost. In order to maintain the uniqueness of this

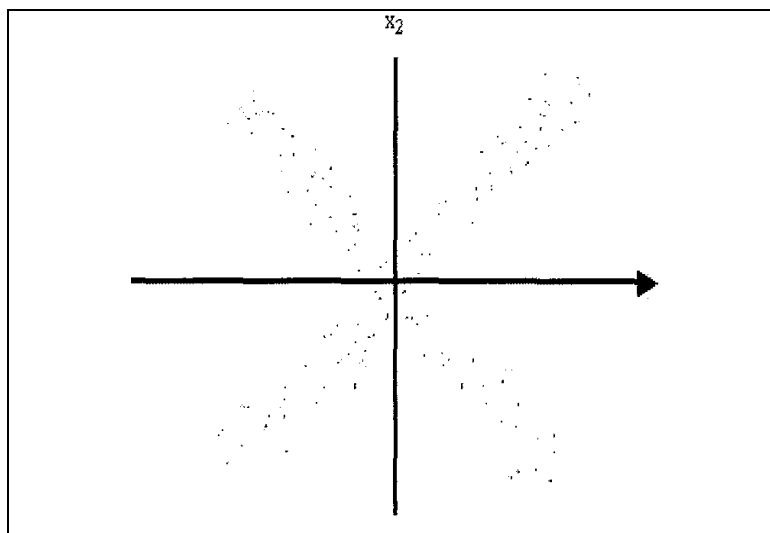
technique, only several Principal Components will be used to represent most of the original data variation. If  $a$  Principal Components are decided to retained with  $a < m$ , then Equation 2.9 can be written as follow:

$$\mathbf{X} = \mathbf{P}_a^T \mathbf{V}_a + \mathbf{P}_{m-a}^T \mathbf{V}_{m-a} \quad (2.13)$$

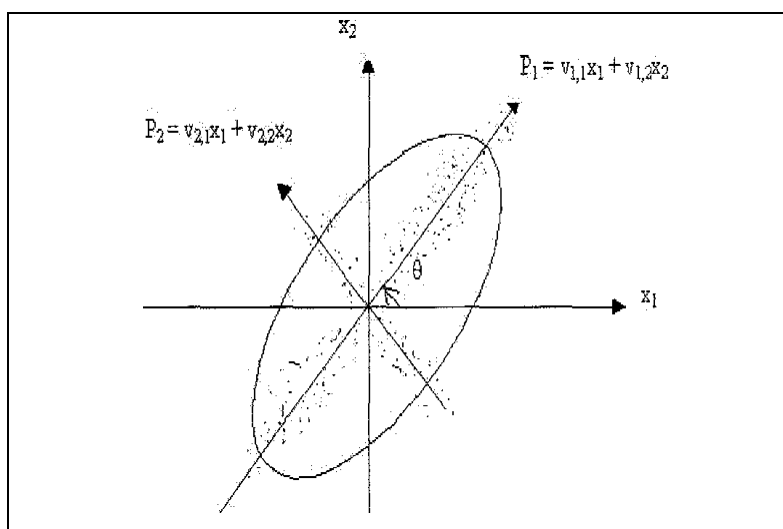
$$\mathbf{X} = \mathbf{P}_a^T \mathbf{V}_a + \mathbf{E} \quad (2.14)$$

The retained principal components  $[P_1, P_2, \dots, P_a]$ , which form the  $P; V_a$  term, are associated with systematic variation in data while the residual principal components  $[P_{a+1}, P_{a+2}, \dots, P_m]$ , which form the residual matrix  $\mathbf{E}$  are considered of containing measurement errors (Seborg *et al.*, 1996). Therefore, PCA is a multivariate analysis technique that could use less number of newly formed variables to represent the original data variations without losing significant information. Information here is referred to data variation.

For the graphical representation of PCA, the linear combinations of the original variables in forming the new variables are actually representing selection of a new coordinate system with  $[P_1, P_2, \dots, P_m]$  as the new axes obtained by rotating the original system with  $x_1, x_2, \dots, x_n$  as the coordinate axes. The new axes represent the direction with maximum variability and provide a simpler and more parsimonious description of the variance-covariance matrix or correlation matrix (Johnson and Wichern, 1992). Figure 2.2 and 2.3 are prepared to give graphical representations of PCA.



**Figure 2.2** System for two variables distribution



**Figure 2.3** graphical representations for PCA

For example, Figure 2.2 shows a system has two variables  $x_1$  and  $x_2$ . In order to have a better representation of the system variability in term of linear combinations of  $x_1$  and  $x_2$ , the original axes are rotated to  $\theta$  degree so that the newly formed Principal Component 1 will represent the maximum variation of the system. Maximum variation is achieved if the perpendicular distance of all the points to the line is minimum. The Principal Component 2 will represent the maximum variation of the remaining variation. The process of forming new variables is repeated until all the Principal Components are determined for the system, which has more than two variables. The representation of axes rotation to form Principal Components is presented in Figure 2.3. Jackson (1980) described that the principal axes ( $P_1$  and  $P_2$ ) are the rotation of the covariance matrix or correlation matrix, and the *eigenvectors* (coefficients of principal components) are the direction cosines of the new axes related to the old. The graphical representation for more than three axes is impossible.

### 2.2.1 Original Data Matrix, Variance-Covariance Matrix and Correlation Matrix

The initial step of Principal Components Analysis is to have a set of data and arrange in the original matrix  $\mathbf{X}$ , which has  $n$  measurements for  $m$  variables. Actually, PCA enable the variance-covariance structure or correlation structure of the original data to be explained by forming the several linear combinations of original variables because the *eigenvectors* that used to form the new variables are determined from the variance-covariance matrix or correlation matrix of original variables (Johnson and Wichern, 1992). Variance is a measure of a variable's dispersion or variation while covariance is a measure of co-variation between two variables. The variance and covariance of each variable are determined respectively from Equation 2.12 and 2.13. The correlation  $r_{j,k}$  is a measure of linear association between two variables like variable  $j$  and variable  $k$ . The correlation between variables and correlation matrix are determined from Equation 2.15 and 2.16 accordingly.

$$\text{Variance, } s^2_{j,j} = \frac{\sum_i^n (x_{i,j} - \bar{x}_j)^2}{n-1} \quad (2.15)$$

$$\text{Covariance, } s_{j,k} = \frac{\sum_i^n (x_{i,j} - \bar{x}_j) \sum_k^n (x_{i,k} - \bar{x}_k)}{(n-1)(n-1)} \quad (2.16)$$

$$\text{Variance-covariance matrix, } S = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,m} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,m} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ s_{m,1} & s_{m,2} & \cdots & s_{m,m} \end{bmatrix} \quad (2.17)$$

$$\text{Correlation, } r_{j,k} = \frac{s_{j,k}}{\sqrt{s_{j,j}} \sqrt{s_{k,k}}} \quad (2.18)$$

$$\text{Correlation matrix, } R = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,m} \\ r_{2,1} & r_{2,2} & \dots & r_{2,m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ r_{m,1} & r_{m,2} & \dots & r_{m,m} \end{bmatrix} \quad (2.19)$$

Where:  $i$  : i-th measurement  $\overline{x_k}$  : mean value for variable  $x_j$

$s^2_{j,j}$  : variance of variable  $x_j$   $s_{j,k}$  : covariance of variable  $x_j$  and  $x_k$

The selection whether the variance-covariance matrix S or correlation matrix R to be used for the Principal Components Analysis is mainly based on the nature of the data. If the involved data are all in the same unit, same digit, and same magnitude order or the data are in percentage, and then the analysis can be carried out on variance-covariance matrix. If the data or measurements are in different unit like pressure, temperature, flow rate, energy and so on, or the data are in different magnitude order, then the Principal Components Analysis shall be performed on the correlation matrix (Marriott, 1974).

### 2.2.2 Determine the Eigenvalues and Eigenvector of the Matrix

The PCA is performed either on variance-covariance matrix or correlation matrix. Let A represents either variance-covariance matrix S or correlation matrix R. In order to determine *eigenvalues* and *eigenvectors* of A, the characteristic equation is form based on the Equation 2.17. Equation 2.17 has to follow the criteria as shown in Equation 2.18. However, the function equation can be expressed in the polynomial form as shown in Equation 2.19 and is called the characteristic equation of matrix A:

$$f(\lambda) = |A - \lambda I| = \begin{vmatrix} a_{1,1} - \lambda & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} - \lambda & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} - \lambda \end{vmatrix} \quad (2.20)$$

$$f(\lambda) = |A - \lambda I| = 0 \quad (2.21)$$

$$c_0 \lambda^n + c_1 \lambda^{n-1} + \dots + c_{n-1} \lambda + c_n = 0 \quad (2.22)$$

The order of the characteristic equation equals to the number of variables involved in the Principal Components Analysis. The solutions of the characteristic equation produces *characteristic root*, *latent root*, or *eigenvalues*,  $\lambda_1, \lambda_2, \dots, \lambda_m$ . The *characteristic vectors*, or *eigenvectors* can be obtained from the following equations:

$$|A - \lambda I| w_i = 0 \quad (2.23)$$

$$v_i = \frac{w_i}{\sqrt{w_i' w_i}} \quad (2.24)$$



Vectors  $w_1, w_2, \dots, w_n$  are the un-normalized *eigenvectors*.  $v_1, v_2, \dots, v_m$  are normalized *eigenvectors*.  $V$  is the *eigenvectors* matrix of variance-covariance matrix or correlation matrix. Matrix  $V$  is ortho-normal that is:  $v_i^T v_i = 1$  and  $v_i^T v_k = 0$ . This means that the newly formed variables or Principal Components are independent with each other or in another word, orthogonal.

### 2.2.3 Number of Principal Components Required

*Eigenvalue* is a measure of the variation explained by an *eigenvector*. Each *eigenvector*,  $v_i$  associates with its *eigenvalue*,  $\lambda_i$ . The first Principal Component explains the largest variation of the original data. The  $m^{\text{th}}$ -Principal Component explains the smallest variation. If the variance-covariance matrix is used for Principal Components Analysis, the total *eigenvalues* equals to the total variance of the original data matrix. If the correlation matrix is used, the total *eigenvalues* equals to the number of the variables involved because the variance for each variable equals to one after standardization. The percentage of the variation to be explained by a particular *eigenvector* is determined via Equation 2.22 while the accumulated variation to be explained until  $j^{\text{th}}$  Principal Component is calculated via Equation 2.23:

Percentage of original variation explained by principal component  $i$

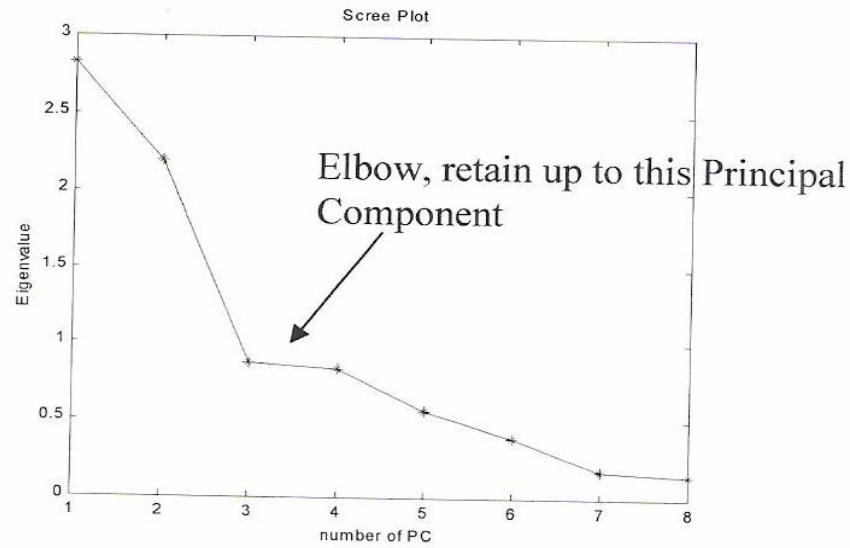
$$= \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_m} \times 100\% \quad (2.25)$$

Percentage of original variation explained by first  $j$  principal components

$$= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_j + \dots + \lambda_m} \quad (2.26)$$

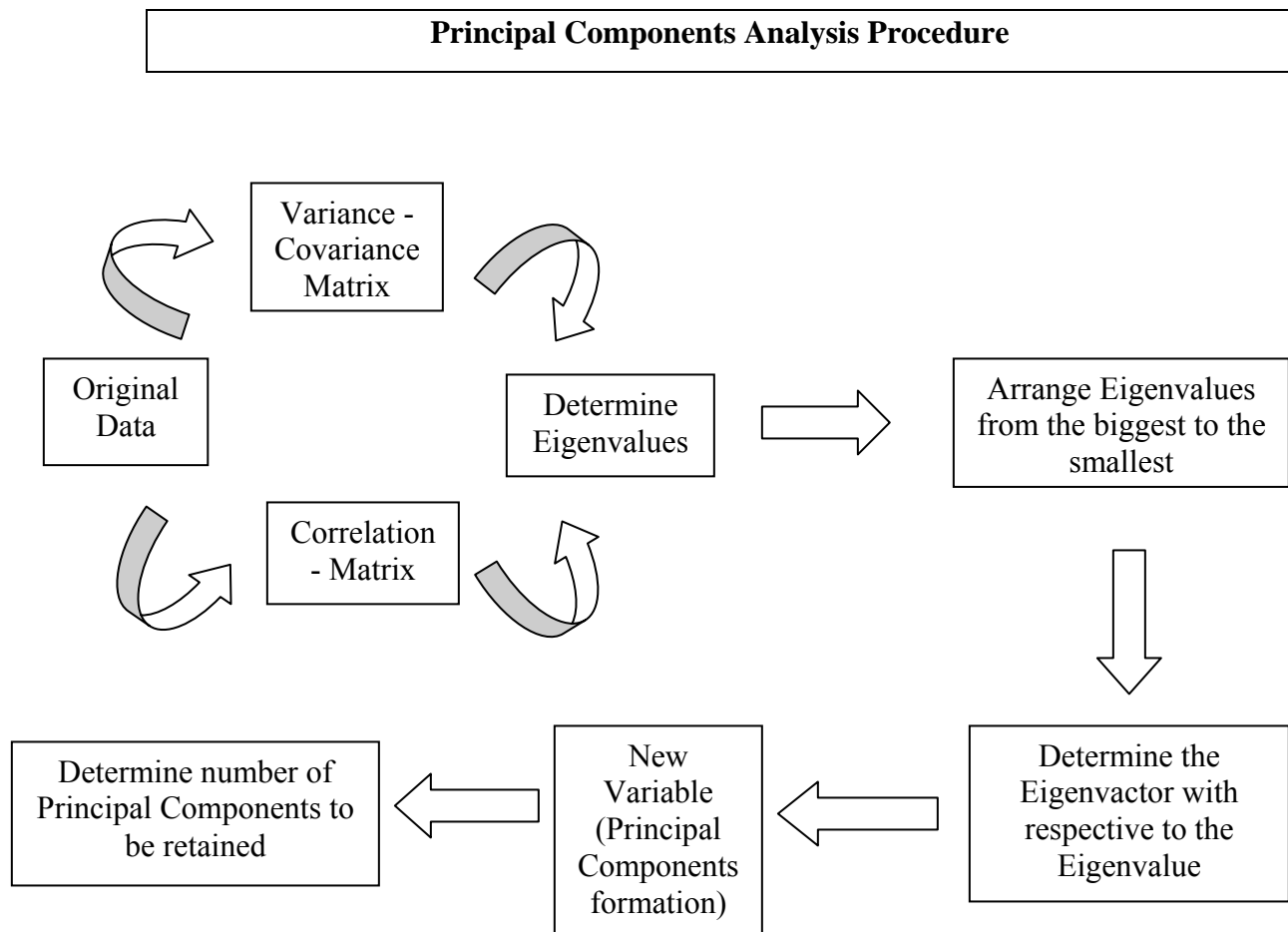
When PCA is applied to reduce the original variables' dimensions, the percentage of the variation to be explained by the Principal Components can be fixed by the user. However, there are several suggestions on determining the number of Principal Components to be retained (Sharma, 1996):

- i) In the case of standardized data, retain only those components whose *eigenvalues* are greater than one. This is referred to as the *eigenvalues-greater-than-one* rule.
- ii) Plot the percent of the variance accounted by each principal component and look for an elbow. The plot is referred to scree plot, which is shown in Figure 2.4.
- iii) Retain components that are statistically significant



**Figure 2.4** Scree Plot

If more than half numbers of total Principal Components are required to account for a reasonable variation, then the data cannot be summarized very well with a few components and shall stick to the original data itself (Cliff, 1987). The summary procedures of Principal Components Analysis are presented in Figure 2.5.



**Figure 2.5** Principal Component Analysis Procedures

## **CHAPTER III**

### **LITERATURE REVIEW**

#### **3.1 Gas Density Measurement**

E. Sanwani presents and interprets results of experimental measurements of the spatial gas hold-up distribution in a 3 m<sup>3</sup> glass rectangular flotation cell at the JKMRC using two different techniques. The gas hold-up device with the capturing technique was developed at the JKMRC and has been used widely in the P9 project1 while the one with conductivity technique was developed at the CSIRO Thermal and Fluids Engineering laboratory at Highett, Victoria, Australia. Measurements were conducted at more than 64 locations in the cell to determine the local gas hold-up distribution in the cell. Since the measurements using the two techniques were conducted at the same locations, the results may be compared with each other. The results indicate that the gas hold-up varies widely inside the flotation cell. The gas hold-up distributions measured by the two techniques are relatively similar except in some locations which can be reasonably explained. [3].

On the base of modern probability approached by the Derevich I.V., the theoretical model of turbulent relative motion of particles in the turbulent flow is developed. Closed equation for probability density function of coordinates and velocities of two particles in turbulent flow is obtained. The system of equations for balance of mass, averaged velocities and intensities of turbulent chaotic motion of particles with account of correlated motion of particles are deduced. The closed expressions for intensity of relative chaotic motion between particles are obtained on